
Title	Is English-medium instruction effective in improving Chinese undergraduate students' English competence?
Author(s)	Jun Lei and Guangwei Hu
Source	<i>International Review of Applied Linguistics in Language Teaching</i> , 52(2), 99–126; doi: 10.1515/iral-2014-0005
Published by	Walter de Gruyter

This document may be used for private study or research purpose only. This document or any part of it may not be duplicated and/or distributed without permission of the copyright owner.

The Singapore Copyright Act applies to the use of this document.

The following article was published originally in Lei, J., & Hu, G. (2014). Is English-medium instruction effective in improving Chinese undergraduate students' English competence? *International Review of Applied Linguistics in Language Teaching*, 52(2), 99–126. doi: 10.1515/iral-2014-0005 and may be found at <http://dx.doi.org/10.1515/iral-2014-0005>

The final publication is available at www.degruyter.com

Jun Lei and Guangwei Hu

Is English-medium instruction effective in improving Chinese undergraduate students' English competence?

Abstract: This study investigates whether English-medium instruction (EMI) has an impact on Chinese undergraduates' English proficiency and affect in English learning and use. A cross-section of 136 sophomores and juniors were drawn from an English-medium and a parallel Chinese-medium program. Data included: (a) participants' scores on two national standardized English proficiency tests; (b) their English-related affect as measured by three scales adapted from Gardner's (2004) Attitude/Motivation Test Battery; (c) their perceptions of EMI in Chinese tertiary education elicited with a survey developed by the Chinese Ministry of Education (2006), and (d) interviews with 10 focal students from the English- and Chinese-medium programs. Results showed no statistically significant effect of medium of instruction on English proficiency or affect in English learning and use. However, extent of satisfaction with EMI, perceived necessity for EMI, and perceived increases in study burden had statistically significant effects on the outcome measures. Additionally, prior English proficiency was the strongest predictor of subsequent English proficiency and English-related affect. These findings raise concerns about the quality of the focal English-medium program and point to students' perceptions of EMI and prior English proficiency as crucial influences on further language learning and use.

Keywords: affect in English learning and use, Chinese higher education, English-medium instruction, English proficiency, medium of instruction policy

DOI 10.1515/iral-2014-0005

1 Introduction

Globalization has fuelled an ever accelerating spread of English as an instructional language into institutions of higher learning around the world (Coleman 2006; Graddol, 2006). As part of its ambitious development agenda, China has

Jun Lei: Nanyang Technological University. E-mail: rayjun.lei@gmail.com

Guangwei Hu: Nanyang Technological University. E-mail: guangwei.hu@nie.edu.sg

been actively promoting EMI in its top-tier universities in the past decade. Although by the end of the 20th century English had been firmly established as a core curricular subject from junior secondary school to higher education in China, there was growing dissatisfaction with the quality of instruction found in the traditional English classroom (Hu 2007). Mounting criticism was leveled at the educational system's failure to meet the anticipated demands for advanced English proficiency arising from China's deepening integration into the world system (Hu & McKay 2012; Zhang 2002). As a response, the Ministry of Education (MOE) (2001) started to promulgate EMI as one of its 12 key guidelines for improving the quality of undergraduate education at the beginning of this century. Specifically, it required that within three years 5–10% of undergraduate specialization courses offered by institutions of higher learning across mainland China be taught through English or another foreign language. This requirement was instigated as a strategic move to achieve the goal of “orienting education toward modernization, the outside world, and the future” and to “meet the challenge of economic globalization and technological revolution” (MOE 2001).¹ EMI was envisioned to be capable of developing an international perspective in Chinese students, improving their English proficiency, and providing access to cutting-edge knowledge available in English. To ensure that the universities would take the English-medium requirement seriously, the MOE made the number of English-medium (EM) courses offered an important criterion in higher education assessment (Hu & McKay 2012). In this connection, the Director of the MOE Higher Education Department stated that “the English proficiency of faculty and students is an important indicator of a university's competitiveness and quality of education” (Zhang 2002, p. 5).

As Hu and McKay (2012) noted, the promotion of EMI has been embraced enthusiastically by universities across China. Most Chinese universities have subscribed to the MOE's vision of EMI as a crucial means of bringing higher education in China up to the level of developed countries and strengthening China's competitiveness in the global arena. In addition, they have seen the adoption of EMI as an opportunity to acquire prestige for their programs, raise their national and international rankings, attract more students, and/or improve their graduates' competitiveness in the job market (Hu 2007). Consequently, many universities have used various incentives (e.g., overseas training, course subsidies, salary increases, and favorable workload calculation) to encourage faculty to teach academic subjects in English (Hu 2007; Tong & Shi 2012). According to Pan (2007), for

¹ All quotations from Chinese-language sources and the participants in this study are translated into English by the authors.

example, Fudan University offered its faculty a bonus of 2,000–8,000 RMB yuan for teaching an EM course, and Shanghai University calculated the workload of a course taught entirely in English three times that of an equivalent course taught in Chinese. As a result of such national and institutional policies, EMI has mushroomed in Chinese higher education over the past decade (Hu 2007; Pan 2007). In a survey study of 135 institutions of higher learning across the country, for example, Wu et al. (2010) found that 132 of them had offered EM courses or programs by 2006, with each institution offering an average of 44 courses.

The literature on EMI in Chinese higher education indicates several common trends. First, EM courses are typically taught by relatively young faculty members who have comparatively good English proficiency and/or overseas educational experience (Hu 2007). Second, textbooks published by Anglo-American universities are usually adopted for EMI, and the relative linguistic demand of the available textbooks is often the most important criterion for choosing one over the others (Pan 2007). Third, because of the differing English proficiency of faculty and students, actual use of English varies greatly from course to course and university to university (Pan 2007). Instructional time in English ranges from most of the class time to hardly any use of English in class, with EMI being restricted to the adoption of textbooks written in English. Fourth, there are widespread perceptions of students' low English proficiency as an impediment to their content learning (Wu et al. 2010). Even in EM programs run by top universities, professors often have to provide Chinese summaries of content presented in English to facilitate students' comprehension of curricular topics (Tong & Shi 2012). These trends raise questions about the quality of EMI at Chinese universities and its effectiveness in achieving the envisioned goals. However, there is a paucity of empirical research on the effectiveness of EMI in improving students' English competence in China. In a research synthesis on EMI in China, Zhu and Yu (2010) found over 90 papers published between 2000 and 2010 but most of these publications were either theoretical discussions about EMI or descriptions of EM course/program characteristics. Virtually no empirical investigations into the effect of EMI on students' English proficiency were reported. Thus, after a decade of top-down and bottom-up promotion of EMI, little is known about whether this form of language provision is effective in improving Chinese university students' command of English.

In contrast to the paucity of empirical research on EMI in China, there is an extensive international literature on the use of a target language as a medium of instruction, including French immersion (FI) in Canada, content-based instruction (CBI) in North America, and content-and-language integrated learning (CLIL) in Europe (see Coleman 2006; Stoller 2004; Swain 2000). This literature abounds with theoretical support for integrating foreign language learning with content

learning (Dalton-Puffer 2011; Swain 2000). Specifically, compared with traditional approaches to foreign language education, integrating language learning with content learning can provide a meaningful and low-anxiety context rich in opportunity for language use, which is believed to facilitate the development of target language proficiency and foster positive affect in language learning and use (Dalton-Puffer 2011; Dupuy 2000). Previous studies in Western contexts generally revealed positive effects of integrating foreign language learning with content learning in these respects (Coyle, Hood, & Marsh 2010; Lazaruk 2007). In her comprehensive review of research on CBI, for example, Dupuy (2000) concluded that students in CBI programs not only “make language gains equal or superior to those of students in traditional language classrooms, and at a much faster pace” but also “develop more positive attitudes toward the target language, show increased self-confidence in their ability to use the target language, and express an interest in pursuing its study” (p. 219). Likewise, in their evaluation of the impact of CLIL programs, Coyle et al. (2010) found generally positive evidence of the overall effectiveness of CLIL in improving students’ language proficiency and affect.

Research on CLIL has examined its effect on various aspects of language competence, including lexis (Lo & Murphy 2010; Ruiz de Zarobe 2010), morpho-syntax (Lorenzo, Casal, & Moore 2010), speaking (Admiraal, Westhoff, & de Bot 2006; Burger & Chrétien 2001; Ruiz de Zarobe 2008), and writing (Lasagabaster 2008; Whittaker, Llinares, & McCabe 2011). These studies have generally found advantages of CLIL programs over non-CLIL ones. In a study comparing speech production between a CLIL and a non-CLIL group, for example, Ruiz de Zarobe (2008) found that the former outperformed the latter in pronunciation, vocabulary, grammar, fluency, and content. Apart from enhancing language proficiency, integrating foreign language learning with content learning has also been found to foster motivation and interest in language learning (Pavesi, Bertocchi, Hofmannová, & Kazianka 2001) and “nurture a ‘can do’ attitude towards language learning in general” (Marsh 2000, p. 14). Students in CLIL programs have been found to hold more positive attitudes toward language learning (Lasagabaster & Sierra 2009; Seikkula-Leino 2007) and report lower anxiety (Maillat 2010; Nikula 2007) compared with their peers in non-CLIL programs. In their study of students’ attitudes toward English and the two official languages in Spain (i.e., Basque and Spanish), Lasagabaster and Sierra (2009) found that CLIL programs boosted students’ positive attitudes toward language learning in general.

These encouraging findings notwithstanding, several researchers (Bruton 2011a, 2011b; Leung 2005; Tedick & Cammarat 2012) have raised concerns about some of these studies and their findings. Bruton (2011b), for example, drew attention to problems with research design (e.g., lack of pretests) and possible misinterpretations of findings in several studies. Additionally, several researchers (e.g.,

Coyle & Baetens Beardsmore 2007; Dafouz, Núñez, & Sancho 2007; Dupuy 2000; Sert 2008) have pointed out that most extant research on FI, CBI and CLIL focused on primary and secondary contexts, with little attention paid to tertiary education. In regard to this, some researchers (e.g., Dalton-Puffer, Nikula, & Smit 2010; Dupuy 2000; Lasagabaster 2008) have cautioned against extrapolating findings from one context to another, because the effectiveness of integrating foreign language learning with content learning largely hinges on contextual factors. Thus, there is a clear need for more research on the effect of EMI on language learning in diverse contexts. The present study is an effort in this direction by focusing on an EM program at a major university in southwestern China. It addresses the following research questions:

- (1) Does the focal EM program have an effect on students' English proficiency?
- (2) Does the focal EM program foster students' positive affect in English learning and use?

2 Method

2.1 Research context

The focal university started to offer EMI in 2002, in response to the MOE's (2001) mandate on the use of English as an instructional language in 5–10% of undergraduate specialization courses. At the time of data collection, the focal university had 27 schools and departments, and offered 32 undergraduate specialisms. Seven of these specialisms (e.g., Accounting, Business Administration, Financial Management) had EM programs in which specialization courses were taught in English (i.e., EMI as defined in this article). This study focused on the EM program in Business Administration commenced in 2008. Along with the EM program, there was a parallel Chinese-medium (CM) program. While the EM program recruited 1 class of 40 students each year, the parallel CM program recruited 3 classes.

During their freshman year, both the EM and CM groups took a course called College English. The CM students received 4 weekly hours of College English reading instruction and another 2 weekly hours of an extended College English module they chose from those offered (e.g., Public Speaking Skills in English, Business English Writing, Business English Listening, or English Movie Appreciation). In sophomore year, the CM students took an extended College English module (2 weekly hours) of their choice each semester.

Like the CM students, the EM students received 4 weekly hours of College English reading instruction in freshman year. In addition, they received 2 hours

of English listening and 2 hours of English speaking instruction each week, which were developed specially for all the EM students. Unlike the CM students, in their sophomore year, the EM students did not take an extended College English module of their own choice. Instead, they took Public Speaking Skills in English in the first semester and Business English Writing in the second. College English was not offered from junior year onward for either the EM or CM students.

The EM students began to receive EMI in sophomore year and continued to have EMI through the first semester of senior year. During this period, the EM program consisted of 7–9 compulsory specialization courses and 1 elective each semester, but only 2–3 of the compulsory specialization courses (2 hours per week for each one) were taught in English. For example, the EM students had 2 EM courses (i.e., Managerial Economics and Principles of Marketing) in the first sophomore semester and 3 (i.e., Financial Management, International Marketing, and Management Information Systems) in the second one. In the EM courses, English textbooks were used, lectures were delivered in English for varying proportions of class time, and exams were set and taken in English. The professors of the EM courses were typically Chinese who had obtained their PhDs from overseas universities. In the CM program, however, Chinese textbooks were adopted, lectures were conducted completely in Chinese, and exams were written and taken in Chinese. The professors were also Chinese and had typically obtained their advanced degrees from universities in mainland China. The subject content exams for the EM and CM students were not directly comparable because the curricular content covered in the EM and CM programs was different.

2.2 Participants

Sixty-four EM students and 72 CM students participated in the study. They had studied English for at least six years before they entered university. To investigate possible long-term EMI effects, there was a roughly even split between sophomores and juniors for each program. Table 1 presents the distribution of the participants. For logistic concerns, participants were drawn from intact classes. There were 45 male students and 90 female students; one student did not report gender. An independent-samples *t*-test on the National Matriculation English

Table 1: Distribution of participants by medium of instruction and year of study

	EMI	CMI
Year 2 (sophomore)	34	38
Year 3 (junior)	30	34

Test (NMET) scores for a sample of 33 students showed no significant difference ($t[31] = 0.80, p > .05$) between participants from the EM program ($n = 16, M = 127.31, SD = 7.45$) and those from the CM program ($n = 17, M = 124.82, SD = 10.18$).² However, this does not mean that the two groups were comparable at the beginning of sophomore year, when EMI started for students in the EM program. Indeed, after receiving 8 weekly hours of College English for a year, the EM students ($n = 64, M = 540.13, SD = 44.70$) scored significantly higher on a national standardized English proficiency test for undergraduates (i.e., College English Test Band 4 or CET 4) than their CM counterparts ($n = 72, M = 498.82, SD = 49.33$), $t(134) = 5.09, p < .001$. Given this difference in English proficiency, CET 4 scores were entered in statistical analyses to partial out the effects of prior differences in English proficiency between the two groups.

2.3 Measures

This study investigated effects of EMI on students' English proficiency and affect in English learning and use. English proficiency was measured by the national standardized College English Test Band 6 (CET 6). Affect in English learning and use was measured by three scales adapted from Gardner's (2004) Attitude/Motivation Test Battery (AMTB). The predictor variables included students' biographic data (i.e., gender, year of study), prior English proficiency as measured by CET 4, medium of instruction (i.e., EM or CM), and perceptions of EMI as measured by items culled from a national survey developed by the MOE (2006).

2.3.1 CET 6

CET 6 is used to assess the English proficiency of undergraduate students in China to determine if their English proficiency meets learning objectives in the national College English curriculum standards (National College English Testing Commit-

² The MOE allows students in some provinces/autonomous regions/municipalities to take the matriculation English tests developed and administered by educational authorities in these places. Participants in this study came from 29 provinces/autonomous regions/municipalities. Consequently, many participants did not take the NMET, and only 33 had NMET scores. If the NMET and regionally developed matriculation English tests were assumed to be comparable, 97 participants' scores were available for comparison. With this dataset, there was still no significant difference between the EM students ($n = 45, M = 127.47, SD = 9.67$) and their CM counterparts ($n = 52, M = 126.06, SD = 8.56$), $t(95) = 0.76, p > .05$.

tee 2006).³ The test is designed for college students who major in any discipline other than English. Scores are reported on a 220–710 scale. In its current format, CET 6 consists of four sections: listening comprehension (35 minutes; maximum score: 249), reading comprehension (40 minutes; maximum score: 249), error correction (15 minutes; maximum score: 70), and writing/translation (35 minutes; maximum score: 142). In the listening comprehension section, students listen to conversations and short passages to answer multiple-choice questions and complete a gapped dictation. In the reading comprehension section, students read passages of various length and complete multiple-choice questions, a banked cloze, true-or-false questions, short-answer questions, etc. The error correction section requires them to identify and correct 10 grammatical errors embedded in a short passage. Finally, the writing/translation section asks students to write a composition of no less than 150 words in 30 minutes and to translate 5 sentences from Chinese to English in 5 minutes. Undergraduate students are required to take CET 4 before they take CET 6. CET 4 follows the same format as CET 6 but targets a lower level of English proficiency (see Zheng & Cheng 2008, for a review of the two tests). Validation studies (e.g., Yang & Weir 1998) have generally found that both CET 6 and CET 4 provide valid and reliable assessments of Chinese undergraduates' general English proficiency.

2.3.2 English learning and use affect

This instrument consisted of three scales adapted from Gardner's (2004) International AMTB for English as a foreign language (Gardner 2010). The adapted scales included the measures of Attitudes toward English Learning (e.g., "I really enjoy learning English"), English Class Anxiety (e.g., "I get nervous when I am speaking in my English class"), and English Use Anxiety (e.g., "I would feel uncomfortable speaking English anywhere outside the classroom"). Each scale had 10 items (5 positively keyed and 5 negatively keyed), asking participants to indicate their agreement/disagreement with each statement on a 6-point Likert scale ranging from "strongly disagree" (1) to "strongly agree" (6). All the negatively keyed items in each scale were reverse coded. Thus, a higher score on the Attitudes toward English Learning Scale indicated more positive attitudes toward English learning;

³ Although they are linked to the national curriculum standards, CET 4 and CET 6 are standardized proficiency tests rather than achievement tests. They have all the defining characteristics of proficiency tests: They are not limited to any specific language course/program, assess overall competence in English, are summative and norm-referenced, and gate-keep passage into higher levels of education (Brown & Abeywickrama 2010; Yang & Weir 1998).

a higher score on the English Class Anxiety and English Use Anxiety Scales indicated a higher level of English class anxiety and English use anxiety, respectively.

To accommodate our participants, we modified the original scales in several respects. First, we translated the original items from English into Chinese and asked an experienced translator to translate the items backward from Chinese into English. Second, we modified the original English Class Anxiety Scale to accommodate English use in both EM content courses and English classes. For example, the item “I get nervous when I am speaking in my English class” was changed into “I get nervous when I am speaking English in class.” As a result, the scale was renamed the In-Class English Use Anxiety Scale. Likewise, the original English Use Anxiety Scale was renamed the Out-of-Class English Use Anxiety Scale. Third, we made some minor wording changes to fit the items to the context of this study. For example, the item “I would feel comfortable speaking English where both Japanese and English speakers were present” was changed to “I would feel comfortable speaking English where both Chinese and English speakers were present.” The internal consistency estimates were acceptable for all the three scales: $\alpha = .88$ ($n = 128$) for Attitudes toward English Learning; $\alpha = .83$ ($n = 127$) for In-Class English Use Anxiety; and $\alpha = .78$ ($n = 128$) for Out-of-Class English Anxiety.⁴

2.3.3 Perceptions of EMI

The participants’ perceptions of EMI were elicited with questions adapted from the MOE’s (2006) survey instrument. The original instrument consisted of 16 items that examined undergraduates’ perceptions, beliefs, and attitudes about various aspects of EMI. However, only three items from the instrument that were directly relevant to our research questions were reported in this study. The first item probed participants’ perceived satisfaction with EMI. Only the EM students were asked to answer this question, choosing from “very unsatisfactory,” “unsatisfactory,” “satisfactory,” and “very satisfactory.” For this item, the CM students were assigned to the category of “not applicable.” The second item asked the participants to indicate whether they regarded EMI as necessary. The third item elicited the participants’ perceptions regarding whether EMI would increase their study burden. Both items asked the participants to answer on a binary scale of yes or no.

⁴ The sample sizes for the reliability estimates differed from the total number of students participating in this study because the estimates were calculated before missing values were imputed. See 2.4.2 for further information on how missing data were dealt with.

2.4 Data collection and analyses

2.4.1 Data collection and coding

EM and CM students at the focal university took CET 4 at the end of freshman year, after they completed one-year (i.e., 32 weeks) instruction of College English. They took CET 6 at the end of sophomore year, after they received another 32-week College English instruction. Because EMI started from the beginning of sophomore year, the EM students had already received one-year (i.e., 32 weeks) EMI in Business Administration courses by the time they took CET 6.⁵ We collected participants' CET scores from their department, after obtaining consent from both the participants and the department authorities and assuring them of anonymity and confidentiality. We also collected the participants' biographic data, data on their English learning and use affect, and perceptions of EMI with a written questionnaire which took about 20 minutes to finish. Furthermore, 6 EM students and 4 CM students participated in a follow-up interview that probed their perceptions, beliefs, and attitudes about EMI. In this article, we focus on the questionnaire data, though we also supplement the questionnaire-based quantitative results with findings from a qualitative analysis of the interview data.

The following variables were coded for multiple regression analyses (dummy coding for categorical data is presented in parentheses): (a) gender (male = 0; female = 1); (b) year of study (sophomore = 0; junior = 1); (c) perceived necessity for EMI (unnecessary = 0; necessary = 1); (d) perceived study burden related to EMI (no increase in study burden = 0; increase in study burden = 1); (e) medium of instruction (CM = 0; EM = 1); (f) perceived satisfaction with EMI (unsatisfactory vs. satisfactory; very unsatisfactory vs. satisfactory; not applicable vs. satisfactory)⁶; (g) CET 4 score; (h) CET 6 score; (i) attitudes toward English learning (average score of the 10 items on the scale); (j) in-class English use anxiety (average score of the 10 items on the scale); and (k) out-of-class English use anxiety (average score of the 10 items on the scale). Multiple regression analyses were chosen over simple bivariate correlation analyses because the former would allow the simultaneous evaluation of the relationships between multiple predictors and an outcome variable, the statistical controlling of the effect of other predictor variables, the assessment of the relative importance of each predictor to the relation-

⁵ At the time of data collection, the CET 6 scores of 9 EM students and 13 CM students were not available, probably because they had not taken the test yet.

⁶ Because no participant chose "very satisfactory," three dummy variables were created with "satisfactory" as the reference variable.

ship, and the testing of interactions between predictors (Field 2009; Tabachnick & Fidell 2007).

2.4.2 Data screening and statistical analyses

To address the research questions, hierarchical multiple regression analyses were conducted on CET 6, Attitudes toward English Learning, In-Class English Use Anxiety, and Out-of-Class English Use Anxiety, respectively, with gender, year of study, perceived necessity for EMI, perceived study burden related to EMI, perceived satisfaction with EMI, and medium of instruction (i.e., EM or CM) as predictors. Before the analyses, missing data and assumptions of multivariate analysis were examined (Tabachnick & Fidell 2007). Because no participant had more than one missing value on any of the three affect scales, each participant's missing value was replaced with his or her mean for the scale in question. As for the remaining variables, only CET 6 had a sizeable percentage of missing values (22 cases; 16.18%), with 13 cases from the CM program and 9 from the EM program, whose CET 6 scores were not available at the time of data collection. Little's MCAR tests were not statistically significant ($\chi^2 = 6.84$, $df = 4$, $p > .05$), indicating that the data were missing completely at random. Thus, cases with missing data were deleted listwise, which left 104 cases for regression analyses on CET 6 and 125 cases for the three affect measures. Mahalanobis distance tests revealed no outliers ($p < .001$). Medium of instruction was found to be highly correlated with perceived satisfaction with EMI, $r(133) = .93$, $p < .001$. For this reason, two sets of hierarchical multiple regression analyses were conducted, one with medium of instruction as one of the predictors and the other with perceived satisfaction with EMI as a predictor.

The predictors were entered in the following order. In Step 1, gender, year of study, perceived necessity for EMI, and perceived study burden related to EMI were entered as a block for models predicting scores on the affect measures in both sets of analyses. All the variables except year of study were entered as a block for the models predicting CET 6 scores in both sets of analyses because all participants took CET 6 in sophomore year, and there was thus no difference in year of study. In Step 2, CET 4 was entered to gauge the effects of prior English proficiency in both sets of analyses. In Step 3, while medium of instruction was entered to assess its unique contribution to variances in the outcome variables in the first set of analyses, perceived satisfaction with EMI was entered in the second set of analyses. Notably, this variable subsumed medium of instruction. Finally, in Step 4, the CET 4 X medium of instruction interaction was entered in the first set of analyses to determine whether prior English proficiency (measured

by CET 4) moderated the effect of medium of instruction on the outcome variables; likewise, the interaction between CET 4 and perceived satisfaction with EMI was entered in the second set of analyses. Following recommendations for testing interaction effects in multiple regression (Aiken & West 1991), we centered CET 4 values by subtracting the grand mean from each participant's score.

3 Findings

3.1 Results of hierarchical multiple regression analyses

Table 2 presents descriptive statistics for the continuous variables entered into the planned regression analyses. These descriptive statistics were based on the data available from the whole sample (except for CET 6, which had missing data). Table 3 presents results from the first set of hierarchical multiple regression analyses predicting the four outcome variables, in which medium of instruction was entered as one of the predictor variables.

As shown in Table 3, of the four biographic and perception variables entered in Step 1, perceived necessity for EMI was significantly related to attitudes toward English learning ($B = 0.70, \beta = .31, p < .001$); perceived study burden was significantly related to in-class English use anxiety ($B = 0.38, \beta = .21, p < .05$). These results indicated that students who saw EMI as necessary were more positive about English learning than those who regarded it as unnecessary. Furthermore, students who thought EMI increased study burden experienced a higher level of in-class English use anxiety than those who did not think so. However, while perceived necessity for EMI contributed significantly to the model's prediction of attitudes toward English learning, $R^2 = .16, F(4, 120) = 5.58, p < .001$, perceived study burden did not improve the model's prediction of in-class English use anxiety, $R^2 = .05, F(4, 120) = 1.40, p > .05$.

Table 2: Descriptive statistics for continuous variables

Variable	<i>n</i>	<i>M</i>	<i>SD</i>	Range	
				Potential	Actual
CET 4	136	518.26	51.39	220–710	377–627
CET 6	114	450.95	72.07	220–710	294–640
Attitudes	136	4.23	0.93	1–6	1.10–6.00
In-class	136	3.82	0.72	1–6	2.00–5.70
Out-of-class	136	3.73	0.72	1–6	2.00–5.70

Table 3: Hierarchical multiple regression analyses predicting English proficiency and affect measures using medium of instruction as a predictor

Predictor	CET 6		Attitudes toward English learning		In-class English use anxiety		Out-of-class English use anxiety	
	ΔR^2	β	ΔR^2	β	ΔR^2	β	ΔR^2	β
Step 1	.07		.16***		.05		.05	
Gender		.16		.15		-.00		-.05
Year of study		–		.04		-.01		-.03
Perceived necessity		.11		.31***		-.03		-.09
Perceived study burden		-.13		-.13		.21*		.17
Step 2	.53***		.10***		.09**		.08**	
CET 4		.75***		.33***		-.32**		-.29**
Step 3	.01		.00		.01		.01	
Medium of instruction (MI)		.12		.05		-.08		-.10
Step 4	.00		.00		.01		.01	
CET 4 * MI		.04		-.08		.10		.15
Total R^2	.61***		.26***		.16**		.15*	
n	104		125		125		125	

* $p < .05$; ** $p < .01$; *** $p < .001$.

In Step 2, CET 4 (i.e., prior English proficiency) was significantly related to all the outcome variables. Specifically, it was positively related to CET 6 ($B = 1.09$, $\beta = .75$, $p < .001$) and attitudes toward English learning ($B = 0.01$, $\beta = .33$, $p < .001$) but negatively related to in-class English use anxiety ($B = -0.004$, $\beta = -.32$, $p < .01$) and out-of-class English use anxiety ($B = -0.004$, $\beta = -.29$, $p < .01$). This means that students who scored higher on CET 4 at the end of freshman year also scored higher on CET 6 at the end of sophomore year, held more positive attitudes toward English learning, and experienced lower levels of in-class and out-of-class English use anxiety. Moreover, CET 4 significantly improved the model's prediction of CET 6, $\Delta R^2 = .53$, $\Delta F(1, 99) = 128.50$, $p < .001$; attitudes toward English learning, $\Delta R^2 = .10$, $\Delta F(1, 119) = 15.27$, $p < .001$; in-class English use anxiety, $\Delta R^2 = .09$, $\Delta F(1, 119) = 12.22$, $p < .01$; and out-of-class English use anxiety, $\Delta R^2 = .08$, $\Delta F(1, 119) = 10.30$, $p < .01$. In other words, after controlling for the biographic and perception variables, CET 4 explained 53% of the variance in CET 6, 10% of the variance in attitudes toward English learning; 9% of the variance in in-class English use anxiety; and 8% of the variance in out-of-class English use anxiety.

In Step 3, medium of instruction was also positively related to CET 6 ($B = 18.08$, $\beta = .12$) and attitudes toward English learning ($B = 0.09$, $\beta = .05$) but negatively related to in-class English use anxiety ($B = -0.11$, $\beta = -.08$) and out-of-class English use anxiety ($B = -0.14$, $\beta = -.10$). However, none of these relationships were statistically significant ($p > .05$). Notably, medium of instruction failed to make any unique, significant contribution to the variance in CET 6, $\Delta R^2 = .01$, $\Delta F(1, 98) = 3.03$, $p > .05$; attitudes toward English learning, $\Delta R^2 = .002$, $\Delta F(1, 118) = 0.32$, $p > .05$; in-class English use anxiety, $\Delta R^2 = .01$, $\Delta F(1, 118) = 0.68$, $p > .05$; and out-of-class English use anxiety, $\Delta R^2 = .01$, $\Delta F(1, 118) = 1.05$, $p > .05$. This means that although the EM students on average scored slightly higher on CET 6, held slightly more positive attitudes toward English learning, and experienced slightly lower levels of in-class and out-of-class English use anxiety than their CM peers, none of these differences were statistically significant, after controlling for gender, year of study, the perception variables, and prior differences in English proficiency.

In Step 4, the interaction between CET 4 and medium of instruction was statistically nonsignificant. This means that CET 4 had the same magnitude of effect on students' English proficiency and the affect measures regardless of medium of instruction. After Step 4, the full regression models explained a significant proportion of variance in CET 6, $R^2 = .61$, $F(6, 97) = 24.97$, $p < .001$; attitudes toward English learning, $R^2 = .26$, $F(7, 117) = 5.81$, $p < .001$; in-class English use anxiety, $R^2 = .16$, $F(7, 117) = 2.79$, $p < .01$; and out-of-class English use anxiety, $R^2 = .15$, $F(7, 117) = 2.72$, $p < .05$.

Table 4 presents results from the second set of hierarchical multiple regression analyses where perceived satisfaction with EMI were entered as a predictor

Table 4: Hierarchical multiple regression analyses predicting English proficiency and affect measures using perceived satisfaction with EMI as a predictor

Predictor	CET 6		Attitudes toward English learning		In-class English use anxiety		Out-of-class English use anxiety	
	ΔR^2	β	ΔR^2	β	ΔR^2	β	ΔR^2	β
Step 1	.07		.16***		.05		.05	
Gender		.16		.15		-.00		-.05
Year of study		-		.04		-.01		-.03
Perceived necessity		.11		.31***		-.03		-.09
Perceived study burden		-.13		-.13		.21*		.17
Step 2	.53***		.10***		.09**		.08**	
CET 4		.75***		.33***		-.32**		-.29**
Step 3 ^a	.02		.01		.01		.02	
PS(US vs. S)		-.07		.05		.10		.12
PS(VUS vs. S)		-.09		.09		-.02		.00
PS(NA vs. S)		-.18*		.00		.12		.16
Step 4	.00		.03		.00		.01	
CET 4 * PS(US vs. S)		.06		-.03		-.00		.08
CET 4 * PS(VUS vs. S)		-.04		.16		.03		.07
CET 4 * PS(NA vs. S)		-.04		.13		-.08		-.08
Total R^2	.62***		.30***		.15		.16*	
<i>n</i>	104		125		125		125	

^a PS = Perceived Satisfaction; S = Satisfactory; US = Unsatisfactory; VUS = Very Unsatisfactory; NA = Not Applicable (i.e., CM students).

* $p < .05$; ** $p < .01$; *** $p < .001$.

variable in lieu of medium of instruction. As the same variables were entered in Steps 1 and 2, the results were identical with those obtained in the first set of analyses reported above. However, as shown in Step 3, perceived satisfaction with EMI (“Not Applicable” versus “Satisfactory”) was significantly related to CET 6 ($B = -26.82$, $\beta = -.18$, $p < .01$). This means that the EM students who perceived EMI to be satisfactory made significantly greater gains (i.e., 26.82 on the 220–710 scale) in their English proficiency than the CM students. However, it did not add to the model’s prediction, $\Delta R^2 = .02$, $\Delta F(3, 96) = 1.60$, $p > .05$. Nor was it significantly related to the affect measures.

In the final step, the interaction between CET 4 and perceived satisfaction with EMI was statistically nonsignificant, indicating that prior English proficiency had the same magnitude of effect on students’ subsequent English proficiency and affect in English learning and use regardless of perceived satisfaction with EMI. After Step 4, the full models explained a significant proportion of variance in CET 6, $R^2 = .62$, $F(10, 93) = 15.00$, $p < .001$; attitudes toward English learning, $R^2 = .30$, $F(11, 113) = 4.13$, $p < .001$; and out-of-class English use anxiety, $R^2 = .16$, $F(11, 113) = 1.85$, $p < .05$. The full model for in-class English use anxiety also approached significance, $R^2 = .15$, $F(11, 113) = 1.81$, $p = .06$.

3.2 Results from qualitative analyses of interview data

Our interviews with the six EM students and their four CM counterparts elicited their general language beliefs, views of EMI, perceptions of the focal EM program, language practices related to EMI, and so on. We selected these interviewees using a maximum variation sampling method to map out both diversity and commonality in EMI-related beliefs, views, and practices. The CM students were interviewed individually, and two focus group interviews were conducted respectively with the three sophomores and the three juniors from the EM program. The individual interviews took an average of 30 minutes, and the focus group interviews lasted 54 and 69 minutes, respectively. Both the individual and focus group interviews were conducted in Chinese, audio-recorded, and transcribed verbatim for subsequent qualitative analyses. In our iterative content analyses of the interview data, we borrowed “sensitizing concepts” (Charmaz 2006) from Spolsky’s (2004, 2009) language policy framework that functioned as interpretative devices for identifying themes and issues in terms of language ideology, language management mechanisms, and language practices. This section reports only those themes that are most relevant to the quantitative findings reported in the preceding section, that is, themes concerning expected and actual language practices in the focal EM program that may throw some light on why the program was found to have

no positive effect on either the participants' English proficiency or their attitudes toward English learning, in-class English use anxiety, and out-of-class English use anxiety. All quotations from the interview data are our direct translations from the original.

A recurrent theme emerging from the interviews concerns the perceived gaps between the purported program goals and the enacted practices. For example, when asked for general opinions about the EM program, one EM student (EMS1) complained:⁷

They set fancy goals and made appealing promises for the program. But the implemented program doesn't meet those goals and promises – in terms of the qualifications of the faculty staffing the program, the quality of teaching or attention to students' feedback.

She further explained,

The enacted program is anything but what the [espoused] program promises. The goals and objectives of the program have been much overstated, and many of them are simply “sign-boards” [to attract people]. The program hasn't met its promise of effective EMI, nor has it met our personal expectations.

In her unflattering comments, this student raised issues about various aspects of the focal EM program and explicitly expressed her dissatisfaction with the perceived ineffectiveness of the program. Broad as they were, these comments provided some general hints as to what factors might have contributed to the program's lack of effectiveness in improving the students' English proficiency or fostering positive affect in English learning and use.

There were also many comments on more specific problems which might shed light on the quantitative results reported in the previous section. A common theme manifest in these comments was summarized by EMS1's observation that “EMI surely has problems, and one major problem is our limited English proficiency.” This recognition of students' inadequate English proficiency for EMI was echoed by another interviewee (EMS2):

Speaking English can cause ambiguity and hence misunderstanding. This is particularly true when it involves highly specialized disciplinary content, where many terms may have meanings different from their everyday usages. It is thus difficult for students (myself included) whose English is not good enough to follow the professors in class and, conse-

⁷ To maintain anonymity, a code is used to refer to an interviewee. For example, EMS1 stands for the first interviewee from the EM program, whereas CMS3 refers to the third interviewee from the CM program.

quently, we can easily get confused. I think it's a waste of time. People say that EMI can improve students' English proficiency. I don't think that's true.

Apart from their perceptions of their own inadequate competence to learn through English, the interviewees also reported dissatisfaction with their EM professors' ability to teach in English. EMS3, for instance, opined that although the EM professors were the cream of the faculty, their oral command of English was poor. Some anecdotal evidence in support of EMS3' opinion was provided by CMS3, who told us that one of her friends from the EM program audited a parallel CM course offered for her class purportedly because he had difficulty with his professor's English and could not understand the content of the EM course well enough.

Connected with the limited command of English by the EM professors and students is another prominent theme in our interview data. That is, the professors' and the students' less than adequate English proficiency compelled them to resort to several pedagogical and language practices to mitigate or overcome their language difficulties. Two such practices were repeatedly brought to our attention by the interviewees. One commonly reported practice was code switching to Chinese in and outside of class to "get by." EMS3, for example, informed us that he resorted to Chinese textbooks after class to study the content covered in class. Similarly, CMS3 shared the following observation with us:

According to some EM students I know, they can hardly understand EM lectures. They borrow our Chinese textbooks to study and then just read through their English textbooks when they are preparing for the final exams. They think all they need to do is to recite answers to potential questions in the final exams.

Some students apparently reconciled themselves to the use of Chinese as a coping strategy. EMS2, for one, saw some value in this strategy:

I think Professor X's method really worked. He delivered his lectures in English, but he assigned us a Chinese reference book. We could refer to it when we had difficulty understanding his lectures or the English textbook. . . . So I usually listened to the lectures and then reviewed my lecture notes. When I came across something I didn't understand, I checked up on it in the Chinese reference book, which was an unfortunate coping strategy to digest the content [delivered in English] that I didn't understand.

Other interviewees, however, were rather negative toward this coping strategy. In this regard, EMS1's opinion was representative:

I feel the benefits of EMI are negligible and its drawbacks are rather apparent. After all, there's such a problem – as reported by many, a large number of students cannot under-

stand EMI in the beginning and the professors often resort to Chinese to help them better understand the lectures. But if Chinese is used as a medium of instruction, it makes little sense to have an overseas educated professor to teach us.

The other commonly reported practice adopted by the EM professors to cope with their students' inadequate command of English was to repeat their explanations in the hope that between the repetitions the disciplinary knowledge would sink in. This coping strategy was apparent in the following observation made by EMS1:

For example, Principles of Management, a course I'm taking this semester, involves mathematics. The professor didn't realize, till well into the semester, that we couldn't understand what he was talking about. In particular, when it came to complex mathematic functions, he had to repeat himself, which took a lot of time. Perhaps that wasn't really a waste of time, but he did spend a lot of time explaining some functions and concepts repeatedly. It would probably be easier for him to explain complex functions and concepts in Chinese. And it would also be a lot more convenient for the professor and students to communicate with each other [in Chinese].

EMS3 apparently concurred with EMS1 when he said,

You wouldn't spend so much time looking up new words in the dictionary if you were taught in Chinese. It would also be much easier to understand lectures in Chinese. The professors wouldn't waste time repeating themselves after discovering, halfway into their lectures, that many students didn't understand what they'd been talking about.

Granted that such language practices and coping strategies as described above could go some way toward helping the students understand the content of the EM courses and, consequently, get by, they apparently did little to improve the students' communicative competence in English. This leads us to the final theme emerging from the interview data that we present here; that is, contrary to the envisioned linguistic goals of the EM program, the actual language practices in the classroom did not seem to create a language acquisition-rich environment. Rather, they appeared to bring about the concomitant consequences of reduced exposure to English in naturalistic contexts of use, diminished pressure on students to engage in meaningful interaction and negotiation via the vehicular language, and impoverished opportunities for them to stretch their linguistic resources and produce extended discourse for genuine communication. EMS1 pointed out:

We generally manage to understand textbooks and lectures. But most of us still find it difficult to speak or write in English. For instance, we may understand some specialized terms when we hear or read them, but we may not be able to spell them.

In a similar vein, EMS2 noted:

We don't have many opportunities to communicate with foreigners. Just think about it, most of our professors are Chinese, and our courses are mostly theory-oriented. There's essentially no such thing as real communication We can understand our textbooks. But we really need discussions in class. That's how we can improve our ability to express and articulate ourselves in English. When you go to the West, what they value most is your ability to contribute [speak and write], rather than your ability to read. I think there isn't enough productive use of English in the program. Anyway, that has always been our problem with English learning.

The growing research literature on effective EMI, however, suggests that it is precisely extensive and varied opportunities to engage in genuine interaction, negotiation of meaning, focus-on-form activities, and extended productive use of the target language that facilitate the development of communicative competence (Costa 2012; Hynninen 2012; Smit 2010).

4 Discussion

The results reported above revealed that when the effects of the biographic and perception variables and prior differences in English proficiency were partialled out, the EM students did not outperform their CM peers on CET 6 after receiving EMI for one year. Nor did the EM students have more positive attitudes toward English learning or experience less anxiety when using English in or outside the classroom after receiving EMI for one (sophomores) or two years (juniors). These results suggest that the focal EM program was not effective in improving students' English proficiency and English learning and use affect. Thus, the present study does not corroborate previous findings about positive effects of EMI on English proficiency (e.g., Lasagabaster 2008; Lo & Murphy 2010) and on affect in English learning and use (e.g., Lasagabaster & Sierra 2009; Maillat 2010; Nikula 2007). This inconsistency may be attributed to program and other contextual factors, which, as some researchers (e.g., Dalton-Puffer et al. 2010; Dupuy 2000; Lasagabaster 2008) have pointed out, can impinge on the effectiveness of a particular EM program.

There are several possible explanations of the EM program's observed lack of effects on the students' English proficiency and affect in English learning and use.

First, the amount of EMI offered to the EM students in sophomore year might not have been sufficient to produce observable effects. This explanation, however, is not tenable if we compare the effects of the extra College English instruction and EMI that the EM students received in their freshman and sophomore year. As described earlier, the EM students had 2–3 EM courses each semester in sophomore year, 2 weekly hours per course. In other words, the EM students had 4–6 more weekly hours of classroom exposure to English than their CM peers. This difference in curriculum time for English in sophomore year was substantially greater than that between the EM and CM students in freshman year because the EM students received only 2 extra weekly hours of College English instruction in freshman year. Yet, they made significantly greater gains in English proficiency than their CM peers during their freshman year. They outperformed their CM counterparts on CET 4 given at the end of freshman year, though the two groups did not differ in their English proficiency (as measured by the matriculation English tests) at the beginning of freshman year. This suggests that the amount of instructional time alone is not an adequate explanation of the absence of an EMI effect on students' English proficiency as measured by CET 6.

A more plausible interpretation is that the focal EM program was deficient in important ways so that it was incapable of delivering what is generally expected of EMI – enhancing students' English proficiency and fostering positive affect in English learning and use. This interpretation was supported by the EM students' reported (dis)satisfaction with EMI. Specifically, while 28 of the 64 EM students perceived the EMI to be satisfactory, 28 and 7 students considered it unsatisfactory and very unsatisfactory, respectively.⁸ In addition, the results from the interview data pointed to various problems with the EM program, including gaps between the purported program goals and actual practice, inadequate command of English as the medium of instruction and learning, and poor pedagogical strategies adopted to cope with language difficulties. The extent of perceived (dis)satisfaction with EMI, together with the problems identified in the interview data, raises concerns about the quality of the EM program studied here. They revealed a language learning environment that was lacking many of those contextual and interactional qualities that previous research has identified as facilitative factors for language acquisition in EMI (e.g., Costa 2012; Hyninen 2012; Smit 2010).

The third explanation for the observed lack of effects of the EM program might lie in a possible mismatch between the proficiency tests used in this study and the nature of proficiency targeted and impacted by EMI. As discussed earlier, both

⁸ The numbers do not add up to 64 because one did not respond to this question.

CET 4 and CET 6 are general English proficiency tests. Some researchers (e.g., Gu & Liu 2005; Zheng & Cheng 2008) have questioned the tests' capacity to assess communicative competence because they largely focus on linguistic knowledge. By integrating content and language learning, EMI focuses on communicative use of English in a meaningful, disciplinary context. As a result, EMI might be effective in developing communicative abilities and subject-specific English skills (Coyle et al. 2010; Pavesi et al. 2001; Swain 2000), which CET 6 may have failed to capture.

In contrast to the lack of effects of the EMI, the College English instruction that the EM students received in freshman year was found effective in significantly improving their scores on CET 4. As noted earlier, the EM students received 2 extra hours of English instruction each week in freshman year. This appeared to have contributed to the EM students' greater English proficiency than the CM students as measured by CET 4. Aside from differences in instructional time, there were also differences in the types of College English instruction between the two groups. Specifically, while the CM students received 6 weekly hours of instruction focused largely on reading comprehension, the EM students received 8 hours of instruction, with half of the instructional time devoted to English listening and speaking skills. Although College English instruction is usually geared toward preparation for CET 4 and CET 6 due to the importance placed on these tests by universities and society (Gu & Liu 2005), several EM students gave testimony to the usefulness of the extra speaking and listening instruction by speaking highly of its role in improving their English proficiency. For example, one student explained: "We are usually good at English reading, but we are generally weak at English listening and speaking. So the listening and speaking courses are very helpful." Notably, if we had simply compared the EM and CM students' scores on CET 6 without taking into account their CET 4 scores and without controlling for the effect of prior English proficiency, the effects of the listening and speaking instruction would have been misattributed to the EM program.

Although medium of instruction had no significant effect on English proficiency and the affect measures, the students' perceptions of EMI were found to be significantly related to the two variables. First, perceived satisfaction with EMI was found to be positively related to English proficiency, though it did not improve the overall predictive power of the regression model and thus should be interpreted with caution. While it is possible that perceived satisfaction with the EM program was a function of the students' greater English proficiency in that a threshold level of English proficiency would be necessary to understand EMI (Hu 2008), it is equally, if not more, plausible to argue that only students who were satisfied with the EM program were able to benefit from it in improving their English proficiency because it met their English learning needs. The latter

interpretation suggests an important role of students' perceptions in mediating the effectiveness of EMI. Second, perceived necessity for EMI was significantly and positively related to attitudes toward English learning, which again pointed to a close relationship between the outcome of EMI and students' perceptions. However, our research design did not allow us to determine the causality between these two variables. It could be that perceived necessity for EMI contributed to positive attitudes toward English learning via its expected effect on English proficiency (see Lasagabaster & Sierra 2009). Alternatively, some students might have a strong desire to learn English in the first place, which might in turn lead them to see EMI as necessary to achieve their English learning goals (see Coyle et al. 2010; Dalton-Puffer, Hüttner, Schindelegger, & Smit 2009). Our interview data gave some credence to the second interpretation. Since 9 of the 10 interviewees perceived English learning as a primary goal of EMI, students holding positive attitudes toward English learning are likely to hold positive attitudes toward EMI, too. Third, perceived increase in study burden was significantly and positively related to in-class English use anxiety but not to out-of-class English use anxiety, suggesting that EMI's impact was mainly on in-class rather than out-of-class English use. This may have to do with the fact that China is an EFL or acquisition-poor context where students do not have many opportunities to use English outside the classroom (Hu 2008). The causal direction of the observed relationship, however, could not be determined with our research design. EMI could lead to extra study burden, which would then result in a higher level of anxiety in class, but a reverse relationship is also conceivable. Taken together, these results constituted considerable evidence that students' perceptions of EMI were closely related to its potential to influence English proficiency and affect in English learning and use.

Finally, CET 4 scores emerged as the strongest predictor of all the four outcome variables, accounting for 53% of the variance in CET 6 scores, 10% in attitudes toward English learning, 9% in in-class English use anxiety, and 8% in out-of-class English use anxiety. On the one hand, it was positively related to CET 6 scores and attitudes toward English learning. This is not surprising because initial success can lead to more success and breed confidence and positive attitudes toward learning. On the other hand, CET 4 scores were negatively related to in- and out-of-class English use anxiety. This is also understandable because language use anxiety is conceivably a function of difficulties in communication that stem from a lack of competence in the target language. This is consistent with previous findings (see Gardner 2010; Horwitz 2001) that have consistently noted an inverse relationship between anxiety and language achievement. These results point to the importance of prior English proficiency in the development of further English proficiency and positive affect in English learning and use,

lending support to previous findings (e.g., Cummins 2000; Dupuy 2000; Stryker & Leaver 1997) that students need to reach a threshold level of proficiency in English to benefit from EMI.

5 Conclusion

This study set out to examine whether EMI was effective in improving students' English proficiency and English learning and use affect in an undergraduate EM program at a major university in mainland China. It has yielded several interesting findings. First, the EM program studied here was not effective in improving either students' English proficiency or their English learning and use affect. Second, the intensive English listening and speaking instruction that the EM students received in freshman year appeared to be more effective in improving their English proficiency than the EMI itself. Third, students' perceptions of EMI and prior English proficiency appeared to mediate the effectiveness of EMI in enhancing English proficiency and positive affect in English learning and use. These findings raise serious concerns about the expected effectiveness of EMI in enhancing English proficiency in Chinese higher education. They also point to the influences of such contextual factors as students' perceptions of EMI that nullify absolutist claims about the often optimistically envisioned benefits of EMI.

It is imperative to note that this study has several limitations and thus calls for further research. First, because all participants took CET 6 at the end of sophomore year, we were only able to examine the effects of one-year EMI on students' English proficiency. Future research should examine EMI's effects on English proficiency over a longer period of time, which will enable a more confident conclusion. Second, because of a possible mismatch between the proficiency tests used in this study and the type of language proficiency targeted at in EMI, future research needs to employ proven measures of subject-specific English competence. As Swain (2000) pointed out, "it would seem crucial, if we are to measure the learning that occurs as a result of the research 'treatment,' that we tailor our tests to what happens during that treatment" (p. 206). Third, there is a clear need to study the effects of EMI on subject content learning and on how language and content learning interact with each other because after all the paramount goal of EM programs is that students will learn the required curricular content. If subject content learning suffers as a result of EMI, the *raison d'être* of such instruction is seriously undermined. Finally, although EM courses and programs have mushroomed at universities across mainland China, their quality remains largely unexamined. The findings of this study point to a clear need for policymakers to give careful consideration to empirical evidence concerning whether and to what

extent EMI's envisioned goals are achievable in specific educational contexts before they decide whether EMI should be scaled up or even encouraged.

References

- Admiraal, W., Westhoff, G., and de Bot, K. (2006). Evaluation of bilingual secondary education in the Netherlands: Students' language proficiency in English. *Educational Research and Evaluation*, 12, 75–93.
- Aiken, L. S., and West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.
- Brown, H. D., and Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices* (2nd ed.). White Plains, NY: Pearson Education.
- Bruton, A. (2011a). Are the differences between CLIL and non-CLIL groups in Andalusia due to CLIL? A reply to Lorenzo, Casal and Moore (2010). *Applied Linguistics*, 32, 236–241.
- Bruton, A. (2011b). Is CLIL so beneficial, or just selective? Re-evaluating some of the research. *System*, 39, 523–532.
- Burger, S., and Chrétien, M. (2001). The development of oral production in content-based second language courses at the University of Ottawa. *Canadian Modern Language Review*, 58, 84–102.
- Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. Los Angeles, CA: Sage.
- Coleman, J. A. (2006). English-medium teaching in European higher education. *Language Teaching*, 39, 1–14.
- Costa, F. (2012). Focus on form in ICLHE lectures in Italy: Evidence from English-medium science lectures by native speakers of Italian. *AILA Review*, 25, 30–47.
- Coyle, D., and Baetens Beardsmore, H. (2007). Research on content and language integrated learning. *International Journal of Bilingual Education and Bilingualism*, 10, 541–542.
- Coyle, D., Hood, P., and Marsh, D. (2010). *CLIL: Content and language integrated learning*. Cambridge, UK: Cambridge University Press.
- Cummins, J. (2000). *Language, power and pedagogy: Bilingual children in the crossfire*. Clevedon, UK: Multilingual Matters.
- Dalton-Puffer, C. (2011). Content-and-language integrated learning: From practice to principles? *Annual Review of Applied Linguistics*, 31, 182–204.
- Dalton-Puffer, C., Hüttner, J., Schindelegger, V., and Smit, U. (2009). Technology-geeks speak out: What students think about vocational CLIL? *International CLIL Research Journal*, 1, 18–25.
- Dalton-Puffer, C., Nikula, T., and Smit, U. (Eds.). (2010). *Language use and language learning in CLIL classrooms*. Amsterdam, the Netherlands: John Benjamins.
- Dafouz, E., Núñez, B., and Sancho, C. (2007). Analysing stance in a CLIL university context: Non-native speaker use of personal pronouns and modal verbs. *International Journal of Bilingual Education and Bilingualism*, 10, 647–662.
- Dupuy, B. C. (2000). Content-based instruction: Can it help ease the transition from beginning to advanced foreign language classes? *Foreign Language Annals*, 33, 205–223.
- Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). Thousand Oaks, CA: Sage.

- Gardner, R. C. (2004). *Attitude/Motivation Test Battery: International AMTB research project*. Canada: The University of Western Ontario.
- Gardner, R. C. (2010). *Motivation and second language acquisition: The socio-educational model*. New York: Peter Lang.
- Graddol, D. (2006). *English next: Why global English may mean the end of 'English as a foreign language.'* London: British Council.
- Gu, W., and Liu, J. (2005). Test analysis of college students' communicative competence in English. *Asian EFL Journal*, 7, 118–133.
- Horwitz, E. (2001). Language anxiety and achievement. *Annual Review of Applied Linguistics*, 21, 112–126.
- Hu, G. W. (2007). The juggernaut of Chinese-English bilingual education. In A.W. Feng (ed.), *Bilingual education in China: Practices, policies and concepts* (pp. 94–126). Clevedon, UK: Multilingual Matters.
- Hu, G. W. (2008). The misleading academic discourse on Chinese-English bilingual education in China. *Review of Educational Research*, 78, 195–231.
- Hu, G. W., and McKay, S. L. (2012). English language education in East Asia: Some recent developments. *Journal of Multilingual and Multicultural Development*, 33, 345–362.
- Hyninen, N. (2012). ICL at the micro level: L2 speakers taking on the role of language experts. *AILA Review*, 25, 13–29.
- Lasagabaster, D. (2008). Foreign language competence in content and language integrated learning. *Open Applied Linguistics Journal*, 1, 31–42.
- Lasagabaster, D., and Sierra, J. M. (2009). Language attitudes in CLIL and traditional EFL classes. *International CLIL Research Journal*, 1, 4–17.
- Lazaruk, W. (2007). Linguistic, academic, and cognitive benefits of French immersion. *Canadian Modern Language Review*, 63, 605–628.
- Leung, C. (2005). Language and content in bilingual education. *Linguistics and Education*, 16, 238–252.
- Lo, Y.-Y., and Murphy, V. A. (2010). Vocabulary knowledge and growth in immersion and regular language-learning programmes in Hong Kong. *Language and Education*, 24, 215–238.
- Lorenzo, F., Casal, S., and Moore, P. (2010). The effects of content and language integrated learning in European education: Key findings from the Andalusian bilingual sections evaluation project. *Applied Linguistics*, 31, 418–442.
- Maillat, D. (2010). The pragmatics of L2 in CLIL. In C. Dalton-Puffer, T. Nikula, and U. Smit (eds.), *Language use and language learning in CLIL* (pp. 39–60). Amsterdam, the Netherlands: John Benjamins.
- Marsh, D. (2000) *Using languages to learn and learning to use languages*. Retrieved 3 April 2012 from <http://www.tieclil.org/html/products/pdf/%201%20UK.pdf>.
- Ministry of Education. (2001). *Guanyu jiaqiang gaodeng yuanxiao benke jiaoxue gongzuo tigao jiaoxue zhihang de ruogan yijian* [Guidelines for strengthening undergraduate education and improving the quality of undergraduate programs]. Retrieved 1 August 2011 from <http://www.edu.cn/20030804/3088968.shtml>.
- Ministry of Education. (2006). *Guanyu dui putong gaodeng xuexiao benke zhuanke shuangyu jiaoxue qingkuang jinxing wenjuan diaocha de tongzhi* [Notice on the survey on bilingual instruction in undergraduate content courses]. Retrieved 1 August 2011 from http://www.moe.gov.cn/publicfiles/business/htmlfiles/moe/A08_sjhj/201109/xxgk_124774.html.
- National College English Testing Committee. (2006). *Syllabus for College English Test* (Band 6). Shanghai, China: Shanghai Language Education Press.

- Nikula, T. (2007). Speaking English in Finnish content-based classrooms. *World Englishes*, 26, 206–223.
- Pan, J. Z. (2007). Facts and considerations about bilingual education in Chinese universities. In A.W. Feng (ed.), *Bilingual education in China: Practices, policies and concepts* (pp. 200–215). Clevedon, UK: Multilingual Matters.
- Pavesi, M., Bertocchi, D., Hofmannová, M., and Kazianka, M. (2001). *CLIL guidelines for teachers*. Milan, Italy: TIE CLIL.
- Ruiz de Zarobe, Y. (2008). CLIL and foreign language learning: A longitudinal study in the Basque country. *International CLIL Research Journal*, 1, 60–73.
- Ruiz de Zarobe, Y. (2010). Written production and CLIL: An empirical study. In C. Dalton- Puffer, T. Nikula, and U. Smit (eds.), *Language use and language learning in CLIL classrooms* (pp. 191–212). Amsterdam, the Netherlands: John Benjamins.
- Seikkula-Leino, J. (2007). CLIL learning: Achievement levels and affective factors. *Language and Education*, 21, 328–341.
- Sert, N. (2008). The language of instruction dilemma in the Turkish context. *System*, 36, 156–171.
- Smit, U. (2010). *English as a lingua franca in higher education: A longitudinal study of classroom discourse*. Berlin, Germany: De Gruyter Mouton.
- Spolsky, B. (2004). *Language policy*. Cambridge, UK: Cambridge University Press.
- Spolsky, B. (2009). *Language management*. Cambridge, UK: Cambridge University Press.
- Stoller, F. L. (2004). Content-based instruction: Perspectives on curriculum planning. *Annual Review of Applied Linguistics*, 24, 261–283.
- Stryker, S., and Leaver, B. (1997). *Content-based instruction in foreign language education: Models and methods*. Washington D.C.: Georgetown University Press.
- Swain, M. (2000). French immersion research in Canada: Recent contributions to SLA and applied linguistics. *Annual Review of Applied Linguistics* 20, 199–212.
- Tabachnick, B. G., and Fidell, L. S. (2007). *Using multivariate statistics* (5 ed.). Boston, MA: Pearson/Allyn and Bacon.
- Tedick, D. J., and Cammarata, L. (2012). Content and language integration in K–12 contexts: Student outcomes, teacher practices, and stakeholder perspectives. *Foreign Language Annals*, 45, s28-s53.
- Tong, F. H., and Shi, Q. (2012). Chinese-English bilingual education in China: A case study of college science majors. *International Journal of Bilingual Education and Bilingualism*, 15, 165–182.
- Whittaker, R., Llinares, A., and McCabe, A. (2011). Written discourse development in CLIL at secondary school. *Language Teaching Research*, 15, 343–362.
- Wu, P., Wang, S. G., Jiang, X., Zeng, D. J., Guan, Y. X., and Li, X. F. (2010). *Gaodeng xuexiao shuangyu jiaoxue de xianzhuang yanjiu he shijian tansuo* [An exploratory study of English-medium instruction practices in Chinese higher educational institutions]. Beijing: Higher Education Press.
- Yang, H., and Weir, C. J. (1998). *The CET validation study*. Shanghai, China: Shanghai Foreign Language Education Press.
- Zhang, Y.X. (2002). Jiaqiang shiyongxing yingyu jiaoxue tigao daoxuesheng yingyu zonghe nengli [Strengthening communication-based English language teaching and raising university students' comprehensive English competence]. *Zhongguo Gaodeng Jiaoyu*, 8, 3–6.

- Zheng, Y., and Cheng, L. (2008). Test review: College English Test (CET) in China. *Language Testing*, 25, 408–417.
- Zhu, Y., and Yu, L. (2010). Woguo hanying shuangyu jiaoyu yanjiu xianzhuang [The state-of-the-art of English-Chinese bilingual education in China]. *Foreign Languages and Their Teaching*, 255, 31–36.

Copyright of IRAL: International Review of Applied Linguistics in Language Teaching is the property of De Gruyter and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.